ehp ENVIRONMENTAL HEALTH PERSPECTIVES

ehponline.org

Comparison of Geostatistical Interpolation and Remote Sensing Techniques for Estimating Long-Term Exposure to Ambient $PM_{2.5}$ Concentrations across the Continental United States

Seung-Jae Lee, Marc L. Serre, Aaron van Donkelaar, Randall V. Martin, Richard T. Burnett, Michael Jerrett

http://dx.doi.org/10.1289/ehp.1205006

Online 2 October 2012

NIEHS
National Institute of Environmental Health Sciences

National Institutes of Health
U.S. Department of Health and Human Services

Comparison of Geostatistical Interpolation and Remote Sensing Techniques for Estimating Long-Term Exposure to Ambient $PM_{2.5}$ Concentrations across the Continental United States

Seung-Jae Lee[1], Marc L. Serre[2], Aaron van Donkelaar[3], Randall V. Martin[3,4], Richard T. Burnett[5], Michael Jerrett[6]*

[1] Geospatial Development Department, Risk Management Solutions, Inc., Newark, California, USA

[2] Department of Environmental Sciences and Engineering, School of Public Health, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

[3] Department of Physics and Atmospheric Science, Dalhousie University, Halifax, Nova Scotia, Canada

[4] The Harvard-Smithsonian Center for Astrophysics, Cambridge, Massachusetts, USA

[5] Population Studies Division, Health Canada, Ottawa, Ontario, Canada

[6] Division of Environmental Health Sciences, School of Public Health, University of California, Berkeley, California, USA

*Corresponding author (Michael Jerrett)

Division of Environmental Health Sciences, School of Public Health, University of California, 50 University Hall, Berkeley, CA 94720-7360 USA. Telephone: (510) 642-3960. Fax: (510) 642-5815. E-mail: jerrett@berkeley.edu

Running title: Comparing estimation methods for chronic $PM_{2.5}$ exposure

Key words: Air Pollution; Chronic Exposure; Geostatistics; $PM_{2.5}$; Remote Sensing

Acknowledgments: This research was funded by Health Canada (Grant No. HC-4500209) and the U.S. Centers for Disease Control and Prevention (Grant No. 200-2010-37394). SJL is employed by Geospatial Development Department, Risk Management Solutions, Inc., Newark, California, USA

CFI statement: None of the authors has any actual or potential competing financial interests.

Abbreviations

ACS (American Cancer Society)
AOD (Aerosol Optical Depth)
AQS (Air Quality Subsystem)
CSTM (Composite Space/Time Mean trend model)
FRM (Federal Reference Method)
GAMM (Generalized Additive Mixed Models)
GIS (Geographic Information System)
KC (Kriging with the CSTM)
LUR (Land Use Regression)
MAE (Mean Absolute Error)
ME (Mean Error)
MISR (Multiangle Imaging Spectroradiometer)
MODIS (Moderate Resolution Imaging Spectroradiometer)
MSE (Mean Square Error)
$PM_{2.5}$ (fine Particulate Matter)
RS (Remote Sensing based $PM_{2.5}$ estimates)
SK (Simple Kriging)
SSTM (Separable Space/Time Mean trend model)
S/TRF (Space/Time Random Field)
US EPA (U.S. Environmental Protection Agency)

Abstract

Background: To better understand adverse health effects from chronic exposure to fine particulate matter ($PM_{2.5}$) a need exists to derive accurate estimates of $PM_{2.5}$ variation at fine spatial scales. Remote sensing has emerged as an important means of estimating $PM_{2.5}$ exposures, but there are relatively few studies that compare remote-sensing estimates to those derived from monitor-based data.

Objective: The purpose of this paper is to evaluate and compare the predictive capabilities of remote sensing and geostatistical interpolation.

Methods: We develop a space-time geostatistical kriging model to predict $PM_{2.5}$ over the continental United States and compare resulting predictions to estimates derived from satellite retrievals.

Results: Within about 100 km of a monitoring station, the kriging estimate was more accurate, while the remote sensing estimate was more accurate for locations >100 km from a monitoring station. Based on this finding we developed a hybrid map that combines the kriging and satellite-based $PM_{2.5}$ estimates.

Conclusions: This study is part of a larger investigation aimed at improving the assessment of exposure to ambient air pollution for chronic health effects studies. We evaluated the estimation capability of monitor-based interpolation to monitor-free remote sensing and found that for most of the populated areas of the continental United States, geostatistical interpolation supplied more accurate estimates than remote sensing. The differences between the estimates from the two methods, however, were relatively small. We conclude that in areas with extensive monitoring networks, the interpolation may provide more accurate estimates, but in the many areas of the world without such monitoring, remote sensing can provide useful exposure estimates that perform nearly as well.

Introduction

An extensive body of research has established $PM_{2.5}$ exposure (particles <2.5 μm in aerodynamic diameter) effects on morbidity and mortality (Hu 2009; Laden et al. 2006; Peters 2001; Pope III 2009). Studies using the American Cancer Society (ACS) cohort to assess the relation between particulate air pollution and mortality rank among the most influential and widely cited. Due to this robust association and a lack of other large cohort studies on the long-term effects, the ACS studies have proven important to government regulatory interventions and health burden assessments (Pope III et al. 2004). However, all of the national estimates from the ACS cohort have relied on central monitoring estimates of city-wide PM concentrations, raising the possibility of substantial measurement error.

To better understand adverse health effects from $PM_{2.5}$ a need exists for accurate estimates of the spatiotemporal variation of $PM_{2.5}$ levels at fine space and time scales. Although much of the $PM_{2.5}$ variation is regional due to secondary formation of organic carbon, sulfates and nitrates (Reiss et al. 2007), some $PM_{2.5}$ mass is derived from local combustion, which may lead to variation at finer spatial scales. In some instances, these finer-scale variations in $PM_{2.5}$ have been shown to associate with larger health effects than those that vary regionally (Jerrett et al. 2005), suggesting the potential importance of refining exposure predictions.

There have been several recent attempts to predict PM at a spatial scale finer than observation scales based on land use regression (LUR) models (Moore et al. 2007; Ross et al. 2007), generalized additive mixed models (GAMM) (Yanosky et al. 2008; 2009), hierarchical modeling (Sampson et al. 2011; Szpiro et al. 2010), geostatistical interpolation (Christakos and Serre 2000; Goovaerts et al. 2006; Liao et al. 2006), and remote sensing techniques (Liu et al. 2004; van Donkelaar et al. 2010). These approaches can be classified under two categories; those

involving ground monitor-based estimation (first four approaches above) and those relying on satellite-based (monitor-free) estimation.

To date, only one study has systematically compared monitor-free estimation with an empirical monitor-based approach (Paciorek and Liu 2009). The study used empirical estimates from a Bayesian hierarchical model which employed land use information derived from a geographic information system (GIS). This carefully conducted analysis over the eastern United States demonstrated that when land use and spatial correlations were incorporated into the estimation, there was little additional predictive value from the satellite aerosol optical depth (AOD) retrievals. This insight was based upon investigating the impact of the satellite retrievals on their $PM_{2.5}$ estimation through the use of cross-validation R^2 and corresponding mean squared prediction error. However, Kumar (2010) criticized the study for its inability to distinguish between the natural and anthropocentric sources of $PM_{2.5}$, in part due to uncontrolled meteorological influences.

A recent satellite-based study generated estimates of chronic $PM_{2.5}$ exposure at 10 km gridded locations globally by integrating satellite-derived AOD and a chemical transport model that incorporates meteorology (van Donkelaar et al. 2010). These estimation surfaces depended on remotely sensed data collected during the period 2001-2006. The satellite-based estimates were, however, inevitably influenced by both random and systematic sources of uncertainty associated with AOD retrieval, varying relations between AOD and $PM_{2.5}$, and temporal sampling biases (Hu 2009; Kumar 2010; Paciorek and Liu 2009).

The main goal of the present study is to compare estimates of long-term average $PM_{2.5}$ for the continental U.S. based on a representative geostatistical kriging model (as a purely monitor-based approach using direct $PM_{2.5}$ measurements) with estimates based on remote sensing (as a

monitor-free approach). In doing so, we contribute novel information to the literature by examining the entire continental U.S. rather than limiting the analysis to the eastern portions of the U.S. Our remote sensing model also directly incorporates meteorological estimates into the calculation of $PM_{2.5}$ concentrations. As mentioned, Paciorek and Liu (2009) used a statistical model with auxiliary GIS data input, which is laborious and time-consuming to compile and execute. In contrast, we compared estimates based on remote sensing with those based on monitoring data only to determine the extent to which remote sensing improves estimation.

Our research is part of a larger project to enhance the prediction capabilities of $PM_{2.5}$ at finer spatial resolution over the U.S. and Canada and to conduct detailed assessment of the health effects from particulate air pollution on all-cause and cause-specific mortality based on concentration-response functions from the ACS cohort.

Materials and Methods

Pollution data (Monthly $PM_{2.5}$ data)

We obtained daily $PM_{2.5}$ measurements for the continental U.S. during 1997-2010 (1,742,020 monitor-days) from the Air Quality Subsystem (AQS) of the U.S. Environmental Protection Agency (USEPA). Our analysis was restricted to filter-based monitors using the Federal Reference Method (FRM) - parameter code 88101. The initial daily data were aggregated to obtain monthly averages that reflect seasonal variation in $PM_{2.5}$ (Bell et al. 2007), which reduces the computational burden associated with the use of daily measurement data. Since many monitoring stations were only in service for part of reporting time period, we only included monitoring stations with at least 50% of possible complete samples in a month. Although the USEPA does not provide monthly $PM_{2.5}$ averages on its air pollution data center web sites,

quality assessment was conducted by comparing arithmetic yearly averages based on the monthly data against annual averages for FRM monitors available from the USEPA. The correspondence between USEPA annual averages and annual averages based on the monthly averages was very strong r=0.996. Monthly values were retained for modeling if the data were determined to have the 50% completeness, resulting in monthly data from 1,315 sites for the interpolation method. We selected a random sample of 147 of these sites (Figure 1) for the validation study described below.

Satellite-Based $PM_{2.5}$ Estimates of Long-Term Average (2001-2006)

We obtained 6-year average $PM_{2.5}$ estimates that were derived for a previous study using an integrated remote sensing-chemical transport model approach (van Donkelaar et al. 2010). Ground-level concentrations of $PM_{2.5}$ were estimated using satellite atmospheric composition data combined with local coincident scaling factors from the GEOS-Chem chemical transport model (http://geos-chem.org). Specifically, AOD data from the MODIS (Moderate Resolution Imaging Spectroradiometer) and MISR (Multiangle Imaging Spectroradiometer) satellites were regridded to a 0.1 by 0.1 degree resolution (~10 by 10 km). The AOD retrievals were translated into estimated ground $PM_{2.5}$ using the output from GEOS-Chem simulations. As part of their analysis, van Donkelaar et al. (2010) removed any AOD with an anticipated bias greater than 0.1 or 20% (whichever was larger), and limited the analysis to spatial points with at least 50 acceptable-quality near-daily AOD values. The authors estimated 6-year average exposures in part because satellite information was missing for many spatial points of the 10 km remote sensing grid over time: averaging data over a 6-year period resulted in comprehensive spatial coverage of the satellite AOD data (~95% global coverage) used to derive long-term $PM_{2.5}$

exposure estimates for the 10 km gridded locations. For the present study, we use this estimation surface, hereafter referred to as "RS", as a baseline method for comparison with 6-year average monitor-based $PM_{2.5}$ exposure estimates derived from ground measurements for the same time period.

Both monitor-based (measured) and monitor-free (satellite-based) $PM_{2.5}$ were initially linked to longitude and latitude (in degree), but were thereafter projected to a planar surface (in km) for our analysis.

Kriging

Kriging is a generalized linear regression technique that accounts for spatiotemporal correlations between samples and provides optimal estimates at unmonitored points. The optimal estimates may be obtained by finding weights that minimize the mean square error (Olea 1999). Many linear kriging methods do not integrate information from physical models or higher order statistics regarding non-linearity, non-Gaussianity, or data uncertainty, but kriging is still a useful method to interpolate numerous space/time dynamics. In our analysis we used simple kriging (SK) with a refined smoothing filter (referred to as Composite Space/Time Mean trend model) as described below.

Let us first define the Space/Time Random Field (S/TRF) $Z(\boldsymbol{p})=Z(\boldsymbol{s},t)$ (Christakos 2000) as a random variable (specifically, the $PM_{2.5}$ distribution over space and time) indexed by the two-dimensional spatial location $\boldsymbol{s}$ and the one-dimensional temporal point t, and $Y(\boldsymbol{p})=\log(Z(\boldsymbol{p}))$ as its log transformation. We also define $m_Y(\boldsymbol{p})$ as a deterministic function representing the global mean trends in $Y(\boldsymbol{p})$ that is constructed such that the deterministic transformation $X(\boldsymbol{p})=Y(\boldsymbol{p})-m_Y(\boldsymbol{p})$ produces a homogeneous stationary S/TRF (defined by a locally constant mean

$m_X(\boldsymbol{p})$=E[$X(\boldsymbol{p})$], where E[.] is the expectation operator, and by a covariance $c_X(\boldsymbol{p}, \boldsymbol{p}')$= E[($X(\boldsymbol{p})$-$m_X(\boldsymbol{p})$)($X(\boldsymbol{p}')$-$m_X(\boldsymbol{p}')$)] that is a function of the spatial lag r=||$\boldsymbol{s}$-$\boldsymbol{s}'$|| and temporal lag τ=|t-t'| between points $\boldsymbol{p}$=($\boldsymbol{s}$,t), and $\boldsymbol{p}'$=($\boldsymbol{s}'$,t')). The mean trend function characterizes the systematic trends and spatiotemporal structures of the $PM_{2.5}$ distribution, whereas the covariance function addresses the correlation structures for the S/TRF, taken at a pair of points.

The SK estimation $\hat{\chi}_k$ of $X(\boldsymbol{p})$ at estimation points k is a linear combination of measurements $\boldsymbol{\chi}_d$ (i.e. realization of $X(\boldsymbol{p})$ at data points $\boldsymbol{p}_d$) given by:

$$\hat{\chi}_k = m_X(\boldsymbol{p}_k) + \lambda^T(\boldsymbol{\chi}_d - \boldsymbol{m}_X(\boldsymbol{p}_d)), \quad [1]$$

where λ is a column vector of SK weights (in general the closer composite space/time separation between $\boldsymbol{p}_k$ and $\boldsymbol{p}_d$, the greater the weight), $m_X(\boldsymbol{p}_k)$ is the mean trend of $X(\boldsymbol{p})$ at the estimation point $\boldsymbol{p}_k$, and $\boldsymbol{m}_X(\boldsymbol{p}_d)$ is a column vector of expected values for $X(\boldsymbol{p})$ at the data points. The vector of SK weights is given by (Olea 1999):

$$\lambda^T = \boldsymbol{c}_{k,d}\boldsymbol{c}_{d,d}^{-1}, \quad [2]$$

where $\boldsymbol{c}_{k,d} = \boldsymbol{c}_X(\boldsymbol{p}_k,\boldsymbol{p}_d)$ is a row vector of covariance for $X(\boldsymbol{p})$ between the estimation point and data points, and $\boldsymbol{c}_{d,d} = \boldsymbol{c}_X(\boldsymbol{p}_d,\boldsymbol{p}_d)$ is a covariance matrix for $X(\boldsymbol{p})$ between the data points. Eqs. [1] and [2] are based on so called Ordinary S/TRF that is a limiting case of a more Generalized S/TRF accounting for spatial non-homogeneity and temporal non-isotropy (Christakos 1992).

We implemented space/time SK estimation using the geostatistical library function *BMElib* written in MATLAB (http://www.unc.edu/depts/case/BMELIB, Christakos et al. 2002). *BMElib* provides an extensive suite of computational functions with which to model the space/time global trend $m_Y(\boldsymbol{p})$ and space/time residual covariance $c_X(\boldsymbol{p}, \boldsymbol{p}')$ (see Supplemental Material, Equation [S1]) functions and derive kriging estimates.

Global mean trend models

A way to obtain the global mean trend $m_Y(\boldsymbol{p})=m_Y(\boldsymbol{s},t)$ is to use the Separable Space/Time Mean trend model (SSTM) (Christakos et al. 2002). The SSTM approach first calculates raw spatial means by averaging the measurements at fixed monitoring sites, and raw temporal means by averaging the measurements at fixed monitoring time events. Next, an exponential filter is applied to the raw spatial and temporal means to derive the smoothed spatial mean component $m_Y(\boldsymbol{s})$ and smoothed temporal component $m_Y(t)$, respectively. For example, the smoothed $m_Y(\boldsymbol{s})$ value is calculated for any spatial point $\boldsymbol{s}$ of interest as the weighted average of the raw spatial means, where the weights decrease exponentially with the distance between each $\boldsymbol{s}$ and the location of the monitoring station where that raw spatial mean was calculated. The space/time mean trend, $m_Y(\boldsymbol{s},t)$ is combined as an additive function of $m_Y(\boldsymbol{s})$ and $m_Y(t)$, i.e.:

$$m_Y(\boldsymbol{s},t)=m_Y(\boldsymbol{s})+m_Y(t)-\mu, \qquad [3]$$

where μ is the mean value of $m_Y(t)$, such that $m_Y(t)-\mu$ represents the fluctuation of $m_Y(t)$ around its mean. The $m_Y(\boldsymbol{s})$ denotes persistent spatial characteristics in $PM_{2.5}$ whereas the $m_Y(t)$ captures seasonal trends in $PM_{2.5}$. The mean trend model is "separable" because each of the smoothed components relies on either a purely spatial or purely temporal metric. The SSTM has performed well in numerous smaller-scale (i.e. state- or city-wide) geostatistical studies (Christakos and Serre 2000; Lee et al. 2010; 2011).

A visual inspection of the time series of $PM_{2.5}$ plotted for all monitoring stations (not shown) revealed a 1-year temporal periodicity in $PM_{2.5}$ levels due to seasonal effects. This periodicity was shifted in time depending on where the monitoring station was located. For example, $PM_{2.5}$ levels at a monitoring station in the western U.S. (Figure 2A,B) were highest in November, while

the $PM_{2.5}$ levels at station in the eastern U.S. (Figure 2A,C) were highest in August. We assumed the periodicity could be fit using weights calculated based on an exponentially decaying function. The Composite Space/Time Mean (CSTM) trend model (Figure 2B,C) is based on a composite space/time metrics (neither purely spatial nor purely temporal) and applies an exponential spatial-averaging to the selected measurements to obtain a smoothed mean trend value for each spatiotemporal coordinate $\boldsymbol{p}_j = [\boldsymbol{s}_j\, t_j]$:

$$m_Y(\boldsymbol{s}_j,t_j) = \sum_{i=1}^{N} w_i\, Y(\boldsymbol{s}_i,t_i) \,/\, \sum_{i=1}^{N} w_i, \qquad [4]$$

where $Y(\boldsymbol{s}_i,t_i)$ is the log-$PM_{2.5}$ measurement at point $\boldsymbol{p}_i=[\boldsymbol{s}_i,t_i]$ such that the Euclidean distance between $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$, $d(\boldsymbol{s}_i, \boldsymbol{s}_j) \leq 100$ km, $|t_i - t_j| \leq 12$ months, and the weight w_i is equal to $exp[-d(\boldsymbol{s}_i, \boldsymbol{s}_j)/a_r - |t_i - t_j|/a_t]$, where a_r and a_t are respectively the spatial and temporal ranges of the exponential smoothing function (in our example a_r =50 km and a_t = 3 months).

We used a cross-validation procedure to compare the accuracy of kriging $PM_{2.5}$ estimates based on the CSTM (referred to as KC estimates hereafter) versus the SSTM (i.e., KS estimates) as described in detail in Supplemental Material. We found that the CSTM outperformed the SSTM, and thus we used kriging with the CSTM (KC) to derive our global mean trend $m_Y(\boldsymbol{p})$ and interpolation estimates for comparison with RS-based estimates respectively.

Validation of KC with RS

Since the RS estimates correspond to chronic exposures to $PM_{2.5}$ equivalent to 6-year average values, we compared them to the 6-year average of monthly KC estimates. For validation purposes we removed all 6-year $PM_{2.5}$ averages measured at the 147 randomly selected validation monitors and derived KC estimates based on data from the remaining 1,315 training monitors only. Next we derived KC and RS estimates of the 6-year $PM_{2.5}$ averages for

each of the validation monitors and compared them with the removed measured (true) values to quantify the mapping accuracy of the KC and RS methods. Finally we investigated how mapping accuracy changes for each validation monitor as a function of the distance with its closest neighbor amongst the training monitors. The more details of this validation procedure are as follows:

i. Select one validation site from which to extract monthly measurements between January 2001 and December 2006 and define them as a vector of validation values χ_v;

ii. Use the kriging equation [Eq. 1] with the training dataset to obtain monthly KC estimates χ_k of $PM_{2.5}$ at the validation monitor (so that each value of the vector χ_k is a set of KC estimates of the corresponding vector χ_v);

iii. Calculate the 6-year averages based on the monthly values (i.e. the average $\hat{\chi}_v$ of the χ_v values and the average $\hat{\chi}_{KC}$ of the χ_k values);

iv. Extract the RS estimate $\hat{\chi}_{RS}$ (remote sensing-based estimation surface averaged over the time period 2001-2006) for that validation monitor;

v. Iterate the steps above, choosing another validation site among the 147 sites;

vi. Out of the 147 sets of $\hat{\chi}_{v,}$ $\hat{\chi}_{KC}$ and $\hat{\chi}_{RS}$ values we retained 74 with data that are ≥80% complete (≥58 of 72 possible monthly records available during 2001-2006). This 80% completeness criterion may be customized, but we note that selecting any percent greater than 80 did not substantially alter the validation results;

vii. The 74 validation sites were categorized into 6 groups based on the spatial lag between the validation monitor and its closest training monitor. The length of the first three classes is equidistant (15 km), and thus the upper limit of the third class is about 45 km. Then we were left with the other 15 validation sites to assign: The length of the last three

classes reduces to 10km (fourth lag between 50 and 60 km, fifth 60-70 km, and sixth greater than 70 km though the minimum values for the last class is 90.8 km) to assign a relatively equal number of the remaining validation sites to each of the last three classes;

viii. For each spatial lag class l=1,…,6 and for each method m=KC or RS we calculated the Mean Error ($ME_m^{(l)}$), the Mean Square Error ($MSE_m^{(l)}$) and the Mean Absolute Error ($MAE_m^{(l)}$) (Lee and Wentz 2008). For each spatial lag class we calculate the percent change in MSE and MAE from the RS to KC methods [e.g., % change in $MSE^{(l)}$ = $[(MSE_{KC}^{(l)} - MSE_{RS}^{(l)}) / MSE_{RS}^{(l)}]\times 100$, so that a negative percent change means that KC has a lower estimation error than RS], as well as the correlation coefficients (i.e. Pearson's r and Spearman's ρ) between the validation values $\hat{\chi}_v$ and corresponding estimates $\hat{\chi}_{KC}$ and $\hat{\chi}_{RS}$ within that class lag.

This procedure produces a validation that goes beyond the traditional approach of examining the accuracy of the method across the entire domain with an MSE. Instead we examine the distance away from a monitor at which each method produces a more accurate result.

Results

KC vs. RS

As evidenced by its lower MSE and MAE statistics and negative values for % change in MSE and MAE (Table 1), the KC method outperforms the RS method consistently up to the fifth spatial lag class (corresponding to an average distance of 65.5 km between the estimation point and its nearest measurement site), but conversely the RS method becomes more accurate when the estimation point is about 106 km away from the nearest measurement site. The estimation accuracy of a method along the spatial lags may be affected by (1) spatial distance (between

estimation and data points), and (2) data quality (whether estimates are based on $PM_{2.5}$ measurements or on auxiliary information such as AOD). In the absence of nearby measurements, RS estimation based on local AOD was more accurate than KC estimation based on measurements at a distant monitor.

As seen in Table 1, the MSE and MAE percent change from the RS to KC methods varies gradually across classes. For example the MSE percent change varies gradually from -72.80% in the first class to -55.07% in the fourth class (as opposed to the unstable variation in the absolute value MSE for the KC method, which goes from 1.229 in first class to 0.699 in the fourth class). Focusing on the MSE/MAE percent changes in Table 1, we find that the MSE/MAE percent changes are negative from the first to the fifth classes, while the changes are positive at the sixth class. This indicates that KC performs better than in the first five classes with shorter spatial lags, while the RS performs better in the sixth class with the longest spatial lag.

The KC predictions are positively and strongly correlated with corresponding measurements at any class, as indicated by the *r* values close to 1 (Table 1), with the exception of the *r* value of 0.071 at the fourth class, which is attributed to a small sample size and the presence of an outlier, which once removed results in a recalculated Pearson's *r* of 0.930. Apart from that outlier, KC *r* values are better (closer to 1) than those of the RS method for the first five classes, while the opposite is true for the sixth class. The Spearman's ρ values reveal a similar pattern. This again indicates that KC performs better that RS for short spatial lags, whereas RS becomes the better estimation method at longer distances.

To elucidate the spatial lag at which RS becomes the better estimation method, we plotted the MSE percent change as a function of the spatial lag using the second order polynomial regression that fits the MSE changes in a least-squares sense (R^2=0.9736) (Figure 3). The MSE

percent change is negative (i.e. KC performs better than RS) for separation distances less than 97.8 km, while the opposite is true beyond that separation distance. However, we caution that a different validation dataset or classification may yield different results for the specific distance at which RS performs better than KC.

Implication to mapping

We generated 10 km-gridded estimation points over space by calculating a weighted average of the KC and RS estimates. The weights are negatively related to the MSE of the KC and RS estimates, which vary as a function of the spatial distance between the estimation point and its closest monitoring site. The KC and RS estimates are equally weighted when the separation distance = 97.8 km, the distance at which MSE for the KC and RS estimates are equal (see Results). For a location < 97.8 km from a monitoring site, we set the KC MSE (MSE_{KC}) to 1, and calculate a relative MSE for RS (MSE_{RS}) = 100/(100+q), where q is the negative percent change in the RS MSE relative to the KC MSE (Figure 3). KC and RS weights are respectively defined by:

$$MSE_{RS} /(MSE_{RS} + MSE_{KC}), \quad [5]$$

$$MSE_{KC} /(MSE_{RS} + MSE_{KC}). \quad [6]$$

For any $q \leq -100$ the contribution of RS would be negligible, and we simply set the KC and RS weights to 1 and 0 respectively. For a location ≥ 97.8 km from a monitoring site, we set the MSE_{RS} to 1, and calculate the MSE_{KC} = 100/(100-q), where q is a positive value. KC and RS weights are also based on Eqs. [5] and [6], but they are respectively set to 0 and 1 for any $q \geq 100$.

We generated 10 km-gridded estimation points over space (the grid cell size used by the RS method) using the RS estimates (Figure 4A) and KC estimates (Figure 4B) to calculate a weighted average based on both approaches. The resulting map (Figure 4C) shows estimated $PM_{2.5}$ levels that are higher than those based only on KC estimates in areas with sparse monitoring data in which monitor-based KC estimates may not be accurate.

Discussion

We compared the best available long-term $PM_{2.5}$ estimates using monitor-based methods (KC estimates) and monitor-free methods (RS estimates). The multi-year duration (2001-2006) we depended upon may be interchangeable with the time duration between 2-5 years commonly found in long-term PM exposure-health effect studies. We found a cut-off separation distance of 97.8 km at which the two methods showed an identical estimation performance. $PM_{2.5}$ measurements contributed significantly to the estimation up to a distance of 97.8 km from the measurement site, but the contribution of the measurements to the estimation was negligible beyond that spatial range. Based on the validation results, the KC method is preferable for estimating chronic exposure to $PM_{2.5}$ up to about 100 km from a measurement site, while the RS method performed better beyond that range.

We used a weighted average to combine the KC and RS estimates according to the distance between the measurement point and estimation site, but may refine this approach in the future, for example, by including more monitors from the 1315 training sites for validation purposes (though the complete set of 72 monthly records for 2001-2006 are available from only 160 of the monitors). Moreover the KC method using ground measurements is potentially more accurate than the RS method for deriving estimates over shorter time scales (e.g. yearly, or monthly).

Therefore, developing an efficient way to combine information from the RS estimates and ground measurements (e.g. using the correlation structures between indirect RS data and direct ground measurements rather than simply taking a weighted average of collocated KC and RS estimates) may lead to substantial improvements in the estimation of $PM_{2.5}$ exposure at space/time resolutions of biological relevance for health studies.

The KC method we developed may be useful for a wide variety of human health studies, but the RS method appears to perform better for estimating exposures of populations that live at relatively longer distances from monitors. A significant portion of the U.S. population (according to 2000 statistics from the U.S. Census Bureau) resides near monitors (i.e. 74.2% are within 25 km, 89.8% are within 50 km, 96.5% are within 75 km, and 98.5% are within 97.8 km, the cut-off distance beyond which the RS method was more accurate based on our analysis). Therefore, in jurisdictions with fairly dense monitoring networks such as the U.S., it may be appropriate to assess exposure with the KC method using nearby ground measurements. However, in most other regions of the world where few $PM_{2.5}$ monitors exist, RS provides a critical information source (Brauer et al. 2011). We conclude that because the differences between estimates based on the two methods were relatively small, for the many areas of the world without dense monitoring remote sensing can provide useful exposure estimates that perform nearly as well as ground-based estimates from a dense network.

The ability to estimate $PM_{2.5}$ based on satellite remote sensing of AOD has advanced rapidly in recent years (Hoff and Christopher 2009). Further improvements in accuracy can be expected through advances in retrieval algorithms to infer AOD from measured radiation, improved calculation of the AOD to $PM_{2.5}$ ratio, and new satellite instrumentation. As $PM_{2.5}$ estimation based on remote sensing continues to improve, RS-based estimates may outperform KC

estimates at distances that are closer to monitor locations than the current 97.8km cut-point identified by our analysis. First, additional information concerning land use, traffic, and population may be incorporated to inform $PM_{2.5}$ concentration estimates (Paciorek and Liu 2009). It can process through a multivariate estimation framework. Second it may be possible to use chemical models to derive a more informative covariance structure and thus more accurate interpolation estimates, particularly when $PM_{2.5}$ monitor data are limited. Development of techniques to combine information from remote sensing, models, and monitors should ultimately yield the best estimate of $PM_{2.5}$ distribution.

Conclusions

We developed a geostatistical interpolation method (KC) to estimate chronic exposure to $PM_{2.5}$ over the U.S., and compared these monitor-based estimates with monitor-free RS estimates for constructing an improved assessment of long-term $PM_{2.5}$ exposure. We identified the distance (97.8 km) between estimation sites and monitors within which KC estimates are more accurate than those based on RS, and conversely beyond which RS estimates are superior. This cut-off radius may be used to combine the KC- and RS-based estimates to build an up-to-date map of chronic exposure to $PM_{2.5}$. The exposure map can provide crucial information for improved risk assessment and be used to improve our ability to study associations between long term exposure to air pollution and adverse health effects in the U.S.

REFERENCES

Abrahamowicz M, Schopflocher T, Leffondré K, du Berger R, Krewski D. 2003. Flexible modeling of exposure-response relationship between long-term average levels of particulate air pollution and mortality in the American Cancer Society study. J Toxicol Env Heal A 66:1625-1653.

Brauer M, Amann M, Burnett RT, Cohen A, Dentener F, Ezzati M et al. 2011. Exposure assessment for the estimation of the global burden of disease attributable to outdoor air pollution. Environ. Sci. & Tech, doi:10.1021/es2025752.

Bell ML, Dominici F, Ebisu K, Zeger SL, Samet M. 2007. Spatial and temporal variation in $PM_{2.5}$ chemical composition in the United States for health effects studies. Environ Health Persp 15(7):989-995.

Christakos G. 1992. Random Field Models in Earth Sciences. Mineola, NY:Dover Publications, INC.

Christakos G, Serre M. 2000. BME analysis of spatiotemporal particulate matter distributions in North Carolina. Atmos Environ 34:3393-3406.

Christakos G. 2000. Modern Spatiotemporal Geostatistics. New York, NY:Oxford University Press.

Christakos G, Bogaert P, Serre ML. 2002. Advanced Functions of Temporal GIS. New York, NY:Springer-Verlag.

Goovaerts P, Auchincloss A, Diez-Roux AV. 2006. Performance comparison of spatial and space-time interpolation techniques for prediction of air pollutant concentrations in the Los Angeles area, Society for Mathematical Geology XIth International Congress.

Hoff RM, Christopher SA. 2009. Remote sensing of particulate pollution from space: have we reached the promised land?. J Air & Waste Manage Assoc 59:645-675.

Hu Z. 2009. Spatial analysis of MODIS aerosol optical depth, $PM_{2.5}$, and chronic coronary heart disease. Int J Health Geogr 8(27), doi:10.1186/1476-072X-8-27.

Jerrett M, Burnett RT, Ma R, Pope III CA, Krewski D, Newbold KB et al. 2005. Spatial analysis of air pollution and mortality in Los Angeles. Epidemiology 16:727-736.

Jerrett M, Arian MA, Kanaroglou P, Beckerman B, Crouse D, Gilbert NL et al. 2007. Modeling the intraurban variability of ambient traffic pollution in Toronto, Canada. J Toxicol Env Heal 70(3-4):200-212.

Kumar N. 2010. What can affect AOD-$PM_{2.5}$ association? Environ Health Persp 118(3):A2-3.

Laden F, Schwartz J, Speizer FE, Dockery DW. 2006. Reduction in fine particulate air pollution and mortality: Extended follow-up of the Harvard six cities study. Am J RespirCrit Care Med 173:667-672.

Lee HJ, Liu Y, Coull BA, Schwartz J, Koutrakis P. 2011. A novel calibration approach of MODIS AOD data to predict $PM_{2.5}$ concentrations. Atmos Chem Phys 11:7991-8002, doi:10.5194/acp-11-7991-2011.

Lee SJ, Wentz EA, Gober P. 2010. Space-time forecasting using soft geostatistics: a case study in forecasting municipal water demand for Phoenix, Arizona. Stoch Environ Res Risk Assess 24:283-295.

Lee SJ, Wentz EA. 2008. Applying Bayesian Maximum Entropy to extrapolating local-scale water consumption in Maricopa County, Arizona. Water Resour Res, W01401, doi:10.1029/2007WR006101.

Lee SJ, Bush B, George R. 2011. Analytic science for geospatial and temporal variability in renewable energy: A case study in estimating photovoltaic output in Arizona. Sol Energy 85:1945-1956.

Liao D, Peuquet DJ, Duan Y, Whitsel EA, Dou J, Smith R et al. 2006. GIS approaches for the estimation of residential-level ambient PM concentrations. Environ Health Persp 114(9):1374-1380.

Liu Y, Park RJ, Jacob DJ, Li Q, Kilaru V, Sarnat JA 2004. Mapping annual mean ground-level $PM_{2.5}$ concentrations using Multiangle Imaging Spectrometer aerosol optical thickness over the contiguous United States. J Geophys Res 109, D22206, doi:10.1029/2004JD005025.

Moore DK, Jerrett M, Mack WJ, Künzli NA. 2007. A land use regression model for predicting ambient fine particulate matter across Los Angeles, CA. J Environ Monit 9:246-252.

Olea R. 1999. Goeststatistics for Engineers and Earth Scientists. Boston, MA:Kluwer Academic Publishers.

Paciorek CJ, Liu Y. 2009. Limitations of remotely sensed aerosol as a spatial proxy for fine particulate matter. Environ Health Persp 117(6):904-909.

Peters A, Dockery DW, Muller JE, Mittleman MA. 2001. Increased particulate air pollution and the triggering of myocardial infarction. Circulation 103:2810-2815.

Pope III CA, Burnett RT, Thurston GD, Thun MJ, Calle EE, Krewski D et al. 2004. Cardiovascular mortality and long-term exposure to particulate air pollution: Epidemiological evidence of general pathophysiological pathways of disease. Circulation 109:71-77.

Pope III CA, Ezzati M, Dockery DW. 2009. Fine-particulate air pollution and life expectancy in the United States. New Engl J Med 360:376-386.

Reiss R, Anderson EL, Cross CE, Hidy G, Hoel D, McClellan R et al. 2007. Evidence of health impacts of sulfate- and nitrate-containing particles in ambient air. Inhal Toxicol 19:419-449.

Ross Z, Jerrett M, Ito K, Tempalski B, Thurston GD. 2007. A land use regression model for predicting fine particulate matter concentrations in the New York City region. Atmos Environ 41:2255-2269.

Sampson PD, Szpiro AA, Sheppard L, Lindström J, Kaufman JD. 2011. Pragmatic estimation of a spatio-temporal air quality model with irregular monitoring data. Atmos Environ 45:6593-6606.

Szpiro AA, Sampson PD, Sheppard L, Lumley T, Adar SD, Kaufman JD. 2010. Predicting intra-urban variation in air pollution concentrations with complex spatio-temporal dependencies. Environmetrics 21:606-631.

van Donkelaar A, Martin RV, Park RJ. 2006. Estimating ground-level $PM_{2.5}$ using aerosol depth determined from satellite remote sensing. J Geophys Res 111, D21201;doi:10.1029/2005JD006996.

van Donkelaar A, Martin RV, Brauer M, Kahn R, Levy R, Verduzco C et al. 2010. Global estimates of ambient fine particulate matter concentrations from satellite-based aerosol optical depth: Development and application. Environ Health Persp 118(6):847-855.

Yanosky JD, Paciorek CJ, Schwartz J, Laden F, Puett R, Suh HH. 2008. Spatio-temporal modeling of chronic PM_{10} exposure for the Nurses' Health Study. Atmos Environ 42:4047-4062.

Yanosky JD, Paciorek CJ, Suh HH. 2009. Predicting chronic fine and coarse particulate exposure using spatiotemporal models for the Northeastern and Midwestern United States. Environ Health Persp 117(4):522-529.

No. of validation monitors	**32**	**13**	**14**	**4**	**6**	**5**
Average distance from monitors to estimation sites (km)	**7.6**	**20.9km**	**39.1km**	**56.5km**	**65.5km**	**106.0km**
MSE KC	1.229	1.610	1.871	0.699	1.145	2.762
MSE RS	4.516	5.307	7.320	1.555	3.014	2.230
MSE change (%) RS to KC	-72.796	-69.672	-74.438	-55.066	-62.017	19.270
MAE KC	0.799	1.084	1.172	0.781	0.993	1.377
MAE RS	1.551	1.883	2.264	1.019	1.626	1.279
MAE change (%) RS to KC	-48.512	-42.401	-48.261	-23.384	-38.940	7.223
ME KC	0.228	0.035	0.511	0.108	-0.586	-0.842
ME RS	0.202	0.402	2.264	1.019	0.148	0.088
Pearson *r* KC	0.929	0.873	0.882	0.071	0.861	0.886
Pearson *r* RS	0.733	0.534	0.879	0.644	0.447	0.908
Spearman ρ KC	0.826	0.786	0.917	0.200	0.600	0.800
Spearman ρ RS	0.546	0.615	0.943	0.400	0.600	0.900

Table 1: Validation statistics for the KC and RS methods

MSE: Mean Square Error, MAE: Mean Absolute Error, ME: Mean Error

FIGURE LEGENDS

Figure 1. Monitoring stations for the U.S. Environmental Protection Agency's $PM_{2.5}$ measurements. Training data for estimation were obtained from 1315 sites indicated by circles. Data for validation were obtained from 147 randomly selected validation sites indicated by green triangles).

Figure 2. (A) Map of the U.S. indicating the month of the year when the monthly average $PM_{2.5}$ concentration is highest. (B, C) $PM_{2.5}$ measurements (blue solid line) and corresponding CSTM (red solid line) and SSTM trend (black dotted line) for a single monitoring site in the Western U.S. (B) and Eastern U.S (C). The individual sites are circled in panel A.

Figure 3. Percent change in MSE from the RS to the KC method (triangles) shown as a function of the distance between the validation point and its closest measurement site. The blue curve indicates a second order polynomial regression model that fits the MSE changes.

Figure 4. Average $PM_{2.5}$ exposure estimates at 10-km gridded locations for 2001 - 2006 based on (A) RS (integrated remote sensing-meteorology model), (B) KC, and (C) a combination of the RS and KC methods.

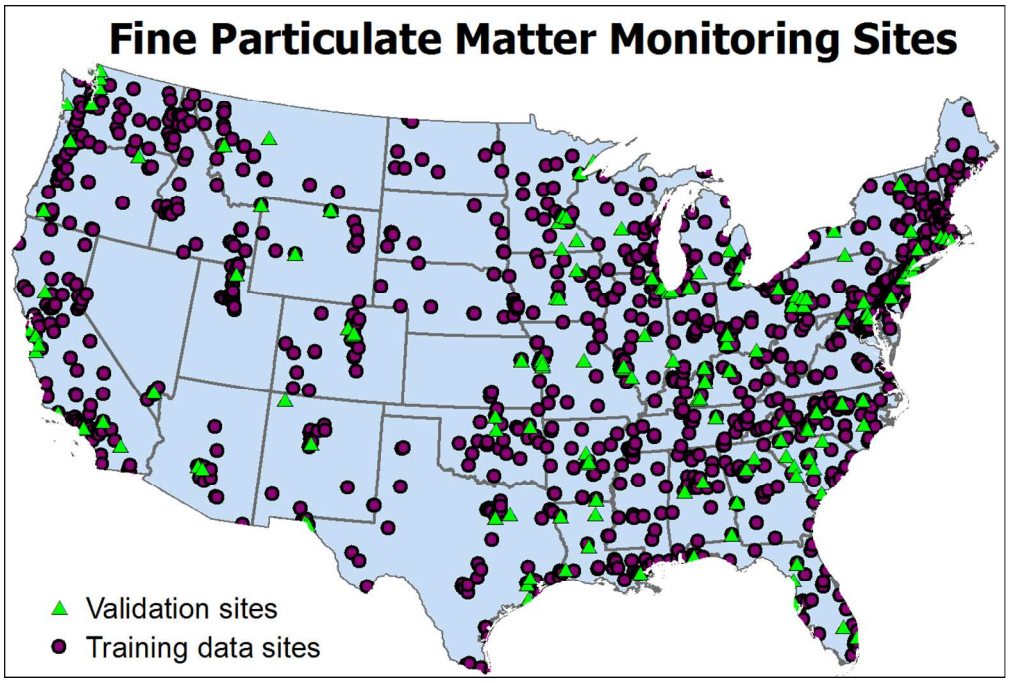

Figure 1. Monitoring stations for the U.S. Environmental Protection Agency's PM2.5 measurements. Training data for estimation were obtained from 1315 sites indicated by circles. Data for validation were obtained from 147 randomly selected validation sites indicated by green triangles).
116x78mm (300 x 300 DPI)

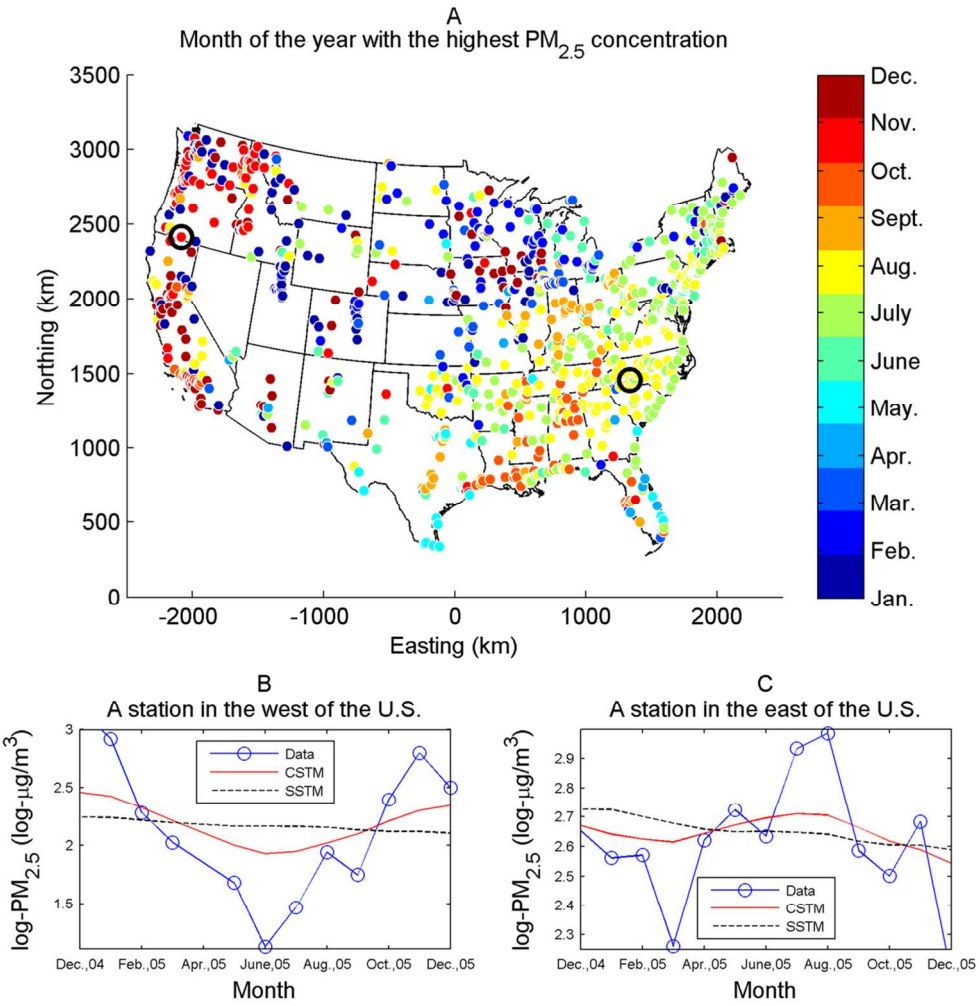

Figure 2. (A) Map of the U.S. indicating the month of the year when the monthly average PM2.5 concentration is highest. (B, C) PM2.5 measurements (blue solid line) and corresponding CSTM (red solid line) and SSTM trend (black dotted line) for a single monitoring site in the Western U.S. (B) and Eastern U.S (C). The individual sites are circled in panel A.
152x152mm (300 x 300 DPI)

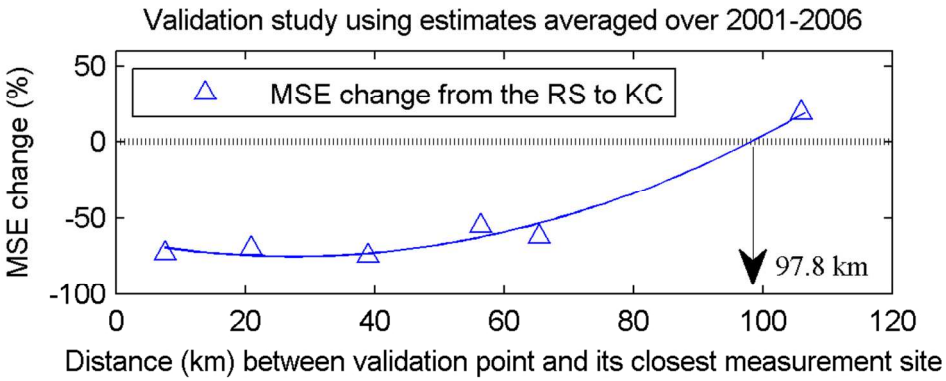

Figure 3. Percent change in MSE from the RS to the KC method (triangles) shown as a function of the distance between the validation point and its closest measurement site. The blue curve indicates a second order polynomial regression model that fits the MSE changes.
101x38mm (300 x 300 DPI)

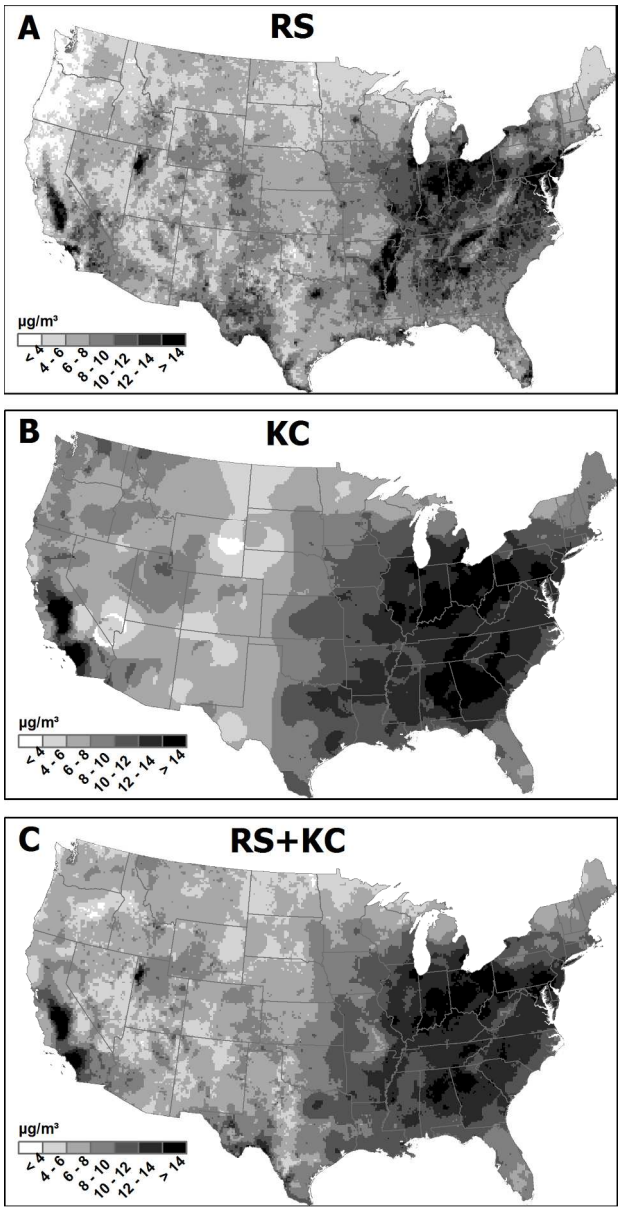

Figure 4. Average PM2.5 exposure estimates at 10-km gridded locations for 2001 - 2006 based on (A) RS (integrated remote sensing-meteorology model), (B) KC, and (C) a combination of the RS and KC methods.
119x233mm (300 x 300 DPI)